Contour Stella Image and Deep Learning for Signal Recognition in the Physical Layer

Yun Lin, *Member, IEEE*, Ya Tu, *Student Member, IEEE*, Zheng Dou, *Member, IEEE*, Lei Chen, *Senior Member, IEEE*, and Shiwen Mao, *Fellow, IEEE*

***Abstract*—The rapid development of communication systems poses unprecedented challenges, e.g., handling exploding wireless signals in a real-time and fine-grained manner. Recent advances in data-driven machine learning algorithms, especially deep learning (DL), show great potential to address the challenges. However, waveforms in the physical layer may not be suitable for the prevalent classical DL models, such as convolution neural network (CNN) and recurrent neural network (RNN), which mainly accept formats of images, time series, and text data in the application layer. Therefore, it is of considerable interest to bridge the gap between signal waveforms to DL amenable data formats. In this article, we develop a framework to transform complex-valued signal waveforms into images with statistical significance, termed contour stellar image (CSI), which can convey deep level statistical information from the raw wireless signal waveforms while being represented in an image data format. In this article, we explore several potential application scenarios and present effective CSI-based solutions to address the signal recognition challenges. Our investigation validates that CSI is a promising method to bridge the gap between signal recognition and DL.**

***Index Terms*—Contour Stella image (CSI), deep learning (DL), signal recognition, physical layer.**

I. Introduction

With the growing diversity and complexity of wireless communication systems, the wireless big data era has arrived [1], [2]. Big data in the signal transmission environment should be extensively studied in order to retrieve interesting and informative information. However, the nature of big data presents great challenges to traditional data analytics algorithms. In particular, data from wireless communications are becoming increasingly complex and heterogeneous, as these are usually collected from various sources with different data formats and complex correlation structures [3], [4]. Moreover, wireless big data can be characterized by four major challenges: massive data, high dimension, rapid change, and quick response requirement [5]–[7]. In the practical world, wireless communication environments are so complex that it is hard to obtain prior knowledge about it. There are many potential benefits of employing deep learning (DL) to address the above challenges: a traditional algorithm feeds all the data into the computer memory and thus has a high storage requirement. Deep neural network (DNN) uses the stochastic gradient descent method to split massive data into small batches and iteratively train the model. Aided by massive data, DL helps to extract complex high-dimension features from wireless signal waveforms, which cannot be precisely modeled via traditional feature engineering approaches [8]. Confronted with rapid changing characteristics of wireless signals, transfer learning can help DNN to quickly adapt to new data before it expires [9]. When it comes to the quick response requirement, Graphics Processing Unit (GPU)-aided parallel computing enables DNN models to make an inference in a short period of time and neural network compression techniques can simplify the model by a large factor without losing too much accuracy. By well analyzing wireless big data via deep learning methods, wireless communication systems can be more effective to provide and support various smart services.

Incorporating the versatile DL into future signal transmission environment systems is drawing increasing interest [10]. For example, Li *et al.* [11] discussed opportunities and challenges of incorporating artificial intelligence (AI) into future network architectures. Peng *et al.* [12] proposed constellation diagrams (CD) to represent wireless signals, which can be used to train a convolutional neural network (CNN) for modulation classification. To achieve quick and precise prediction in complex and heterogeneous wireless communication environments, Zhang *et al.* [9] introduced a deep transfer learning method for intelligent cellular traffic prediction. Tu *et al.* [13] applied semi-supervised generative adversarial networks (SSGAN) to deal with the abundant unlabeled wireless data. In order to deploy DNN into the edge equipment, Tu and Lin [14] introduced a network compression technique for the DNN model.

Furthermore, Tian *et al.* [15] proposed a modulation-constrained (MC) clustering classifier for recognizing the modulation scheme with unknown channel matrix and noise

Manuscript received April 30, 2020; revised August 16, 2020; accepted September 7, 2020. Date of publication September 18, 2020; date of current version March 8, 2021. This work is supported by the National Natural Science Foundation of China (61771154), the Fundamental Research Funds for the Central Universities (3072020CF0813). The associate editor coordinating the review of this article and approving it for publication was G. Ding. *(Corresponding author: Yun Lin.)*

Yun Lin, Ya Tu, and Zheng Dou are with the Key Laboratory of Advanced Marine Communication and Information Technology, College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: linyun@hrbeu.edu.cn).

Lei Chen is with the Department of Information Technology, Georgia Southern University, Statesboro, GA 30460 USA (e-mail: lchen@georgiasouthern.edu).

Shiwen Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: smao@ieee.org).

Digital Object Identifier 10.1109/TCCN.2020.3024610

variance for MIMO systems. The proposed MC classifier together with centroid reconstruction and initialization methods not only reduce the number of parameters to be estimated, but also help to initialize the clustering algorithm for faster convergence. Jiang and Zhu [16] investigated the problem of capacity management in a three-layer heterogeneous satellite network and proposed a low-complexity method for calculating the capacity between satellites, and a long-term optimal capacity allocation algorithm to optimize the long-term utility of the system.

Yu *et al.* [18] proposed a multisampling convolutional neural network (MSCNN) to extract RF fingerprint from the selected region of interest (ROI) for classifying ZigBee devices. A signal-to-noise ratio (SNR) adaptive ROI selection algorithm was also developed to alleviate the effect of the semi-steady behavior of ZigBee devices owing to sleep mode switching. Li *et al.* [17] proposed a system level design of key generation, including quantization, information reconciliation, and privacy amplification. Numerical results verified that the key generation enhanced by PCA with the common eigenvector can high secret key generation rate, low key error rate, and good randomness. Wu *et al.* [19] studied and constructed a robust cognitive user evaluation reference system based on improving the performance of cooperative spectrum sensing. At the same time, the algorithm for the attacker identification and elimination was improved, and the influence of abnormal data on the perceived performance under the combined effect was eliminated. Ding *et al.* [20] proposed a vision, named Dragnet, tailoring the recently emerging Cognitive Internet of Things (IoT) framework for amateur drone surveillance and discussed the key enabling techniques for Dragnet in detail, as well as the technical challenges and open issues. Lin *et al.* [21] proposed an RF fingerprint identification method based on dimension reduction and machine learning as a component of intrusion detection for resolving authentication security issues and obtain the recognition system with the best performance. Lin *et al.* [22] proposed a hybrid spectrum access algorithm based on a reinforcement learning model for the power allocation problem for both the transmission channel and the control channel. The simulation results showed that this new algorithm provides a significant improvement in terms of the trade-off between the control channel reliability and the efficiency of the transmission channel. Liu *et al.* [23] presented a boosting algorithm as an ensemble framework to achieve a higher accuracy than a single classifier. The simulation results showed that gradient boosting had better performance than AdaBoost, and Xgboost achieved the optimal cost performance. Zhang *et al.* [24] proposed a recognition algorithm of D2D devices based on the RF fingerprint and used Hilbert transform and principal component analysis (PCA) to generate the RF fingerprint of D2D devices. Wang *et al.* [25] proposed a low-complexity near-optimal reinforcement learning algorithm for a co-design by considering both complexity and feasibility. Zhang *et al.* [26] studied the problem of interference source identification and obtained a classification accuracy of around 89.5% using any of the four different deep neural network architectures, including CNN, ResNet, CLDNN, and LSTM, which confirmed the feasibility of fast deep learning for wireless interference identification. Dagres *et al.* [27] proposed a platform, which addressed the aspects related to source detection, identification, and localization, and techniques useful in wireless-network coexistence and opportunistic spectrum access. However, relatively few prior studies are concerned with the data conversion method, which brings together the DL and signal recognition fields.

Nowadays, the mainstream data formats for DNN models include image, sequence data, and text data, which are very different from the wireless data that are usually in the time series format. It is of fundamental significance to transform wireless communication waveforms in the physical layer into data formats that are amenable for DNN models to fully utilize the recent advances in deep learning [14]. Under this circumstance, there is a compelling need for a smart way to transform wireless signal waveforms into images while still attaining the statistics information therein. In this article, we propose a contour stellar image (CSI) method based on data density, which can be deemed as a bridge between wireless signal waveforms and DL methods. Combining and transferring DL's classical problems and theory, we explore CSI's potential in connecting the signal recognition field and the DL field. The result shows that CSI is an effective technique to represent signal's statistical information, while its image data format can be directly fed into classical DL models, eliminating the need for customized designs.

The remainder of this article is organized as follows. In Section II, we discuss how to produce CSI from wireless signal waveforms and its merit for signal recognition. In Section III, we provide five potential application scenarios on CSI assisted DL models for signal recognition. In Section IV, we present future research directions. Section V concludes this article.

II. From Constellation Diagram to Contour Stella Image

Generally speaking, information is transmitted as an electromagnetic waveform and captured by the receiver as a series of samples in a digital modulation system that has been well characterized by the CD. Briefly, the rationale of CD is as follows: the signal representing each symbol can be created by adding together different amounts of a cosine wave representing the "I" or in-phase carrier, and a sine wave, shifted by 90 degree from the I carrier and called the "Q" or quadrature carrier. Thus, each symbol can be represented by a complex number, and the CD can be regarded as a set of points (or, dots) in a complex plane, with the horizontal real axis representing the I component and the vertical imaginary axis representing the Q component. The angle of a point (with respect to the origin) represents the phase shift of the carrier wave from a reference phase and the distance of a point from the origin represents the amplitude of the signal. Influenced by channel noise and other types of impairments, the received sample point will be offset from the nominal position and can be clustered into a "cloud" of points surrounding each nominal symbol position in communication scenarios with low SNR.

Based on CD, in this article, we propose a new concept named CSI to fully utilize the dot density concept. The basic

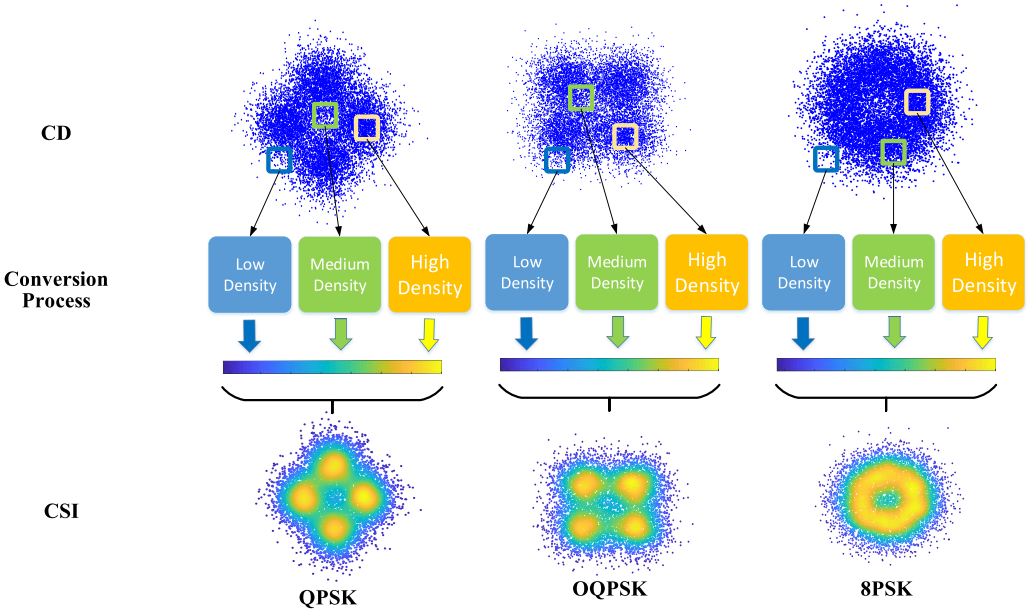


Fig. 1. An illustration of the conversion process from CD to CSI for QPSK, OQPSK, 8PSK when SNR is 4dB. We map CD low density, middle density, and high middle density into three colors: blue, green and yellow.

idea is as follows: to consider the impact of multiple sample points, we choose a window function to slide on the CD and count how many dots are in different regions of the CD. Let us consider a rectangular window function with size $W \times H$, where W is the width and H is the height of the window. We define a metric *dot density* to indicate how many sample dots are located in the window of the complex-valued plane. Let the center coordinate of the rectangular window in the complex plane be (i, j). Then we can normalize the dot counts and obtain the dot density as:

$$\rho_{(W,H)}(i,j) = \frac{\mathrm{dots}_{(W,H)}(i,j)}{N}, \tag{1}$$

where $\mathrm{dots}_{(W,H)}(i,j)$ is the number of dots in this rectangular window, and N denotes the length of the signal I/Q waveform. Next, we map the CD into different colors according to the dot densities of different regions. The entire process is depicted in Fig. 1:

The advantages of the CSI approach are as follows.

(i) Note that a CD is a binary image, which does not distinguish between a pixel with a single sample point and a pixel with multiple sample points. In contrast, the color and shape in CSI will tell more details about the represented wireless signal. In highly noisy communication environments, CSI can provide fine granular features since it considers overlapping sample points. Furthermore, CSI is robust to noise. The color distribution in CSI, such as color and shape of a color zone, can retain the features of the statistics information of signals although under the perturbation of noises.

(ii) CSI retains the statistical information in the signal, including (but not limited to) Gaussian noise, noncoherent single frequency interference, phase noise, amplifier compression, and distortion introduced by the communication environment. In other words, CSI can provide a relatively more comprehensive representation of the wireless communication environment.

(iii) From the perspective of DL, CSI is in a general image format. Classical DL models designed for computer vision (CV), such as AlexNet [29], VGG [30], generative adversarial networks (GANs) [29], and many other DL techniques [8] (e.g., data augmentation [29]), can be directly applied to deal with the CSI data. This will greatly reduce the requirement for the expertise on designing a specific DL model for wireless signals, as well as guaranteeing stable results.

To sum up, CSI is a new class of deep learning-based radio waveform detectors that leverages the powerful new techniques developed in computer vision, especially convolutional feature learning. It holds high potential to improve the signal detection and classification performance of practical systems by generalizing well and remaining sensitive to very low power signals. A strong analogy of this task exists with computer vision with object identification tasks. In doing so, it may finally be possible to develop classifiers that generalize well, are resilient to impairments, and can achieve good classification performance in a wide range of scenarios.

III. Potential Application Scenarios

In this section, we will present five potential application scenarios of the CSI technique. The process on utilizing CSI for DL is illustrated in Fig. 2. As shown in Fig. 2, the general process of converting raw signal waveforms into CSI consists of several steps. First, we collect wireless signal waveforms from wireless emitters at a receiver. Then we split the waveforms into I/Q signal waveforms. After that, we sample the I/Q signal waveforms and map the samples into CSI according to dot density (defined in (1)), which can then be fed into a DL model for training and inference in various application scenarios.

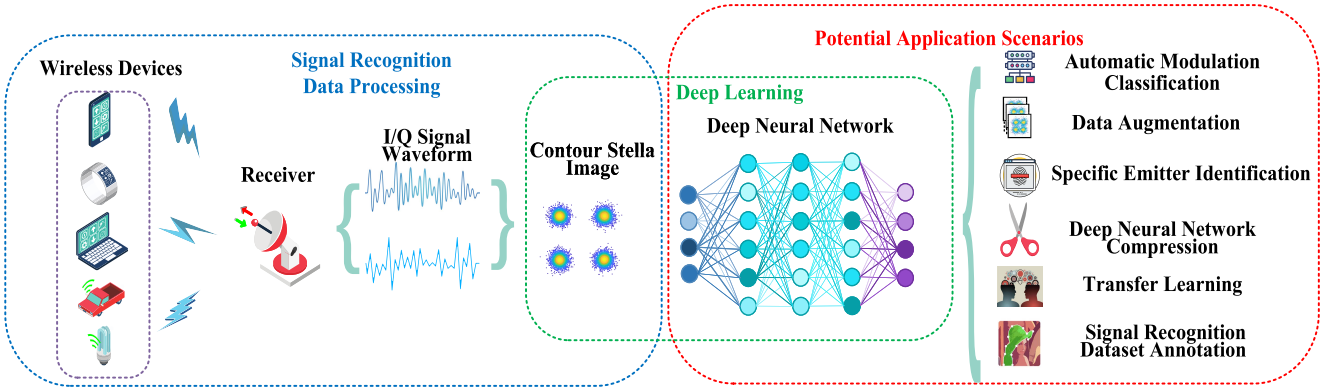


Fig. 2. Illustrate the process of converting wireless signal waveforms into CSI for various application scenarios.

A. Application Scenario I: Automatic Modulation Classification

A signal modulation technique uses a carrier signal to enhance the signal's immunity against channel noise and extend the transmission range. Moreover, modulation can serve as an encryption method and control the bandwidth usage. In order to demodulate the modulated signals to recover the transmitted message, the receiver system needs to know the signal modulation type [31]. Therefore, modulation classification has been widely used in both military and civilian applications. In military applications, modulation classification is useful for recovering the transmitted message and generating jamming signals with matching modulations. In civilian applications, the modulation classification mechanism will help to choose the right demodulation scheme to guarantee recovering the message. This process is initially completed manually by experienced signal engineers and later automated with automatic modulation classification (AMC) systems to improve the overall classification performance [32].

For AMC, CSI inherits the CD features and carries the communication environment statistics information such as Gaussian noise, phase noise, etc. A high performance CNN, like AlexNet, will apply its super non-linear approximation power for achieving close-to-optimal classification accuracy. To prove CSI is a more effective means to learn signal modulation features, we compare our CSI-based approach with the CD color method proposed in [12]. The baseline scheme proposed in [12] combined the impacts of all data samples on each pixel in CD to produce a three-channel enhanced gray image, by considering the distances between sample points and the centroid of pixels,

In the experiment, we create the CSI dataset for semi-supervised learning. The dataset includes 8 categories of modulated signal when the SNR is 4 dB, including BPSK, 4ASK, QPSK, OQPSK, 8PSK, 16QAM, 32QAM, and 64QAM. Each modulated signal category has 10000 labeled data for training and 1000 labeled data for testing, which means there will be 80000 labeled training data and 8000 labeled test data in total. AlexNet and GoogleNet are then applied to the three-channel enhanced gray image dataset. To test the superiority of CSI, we choose the same category of modulated signals but only 1/10 of the training dataset that has been used by the baseline scheme [12]. AlexNet and GoogleNet architecture and modulation signal parameters are the same as in [12]. We also compare our results with a traditional expert feature based scheme. The use of (second- or higher-order) cyclostationarity property of time series has already proved to be very helpful in many applications. One can compute the second, fourth, and sixth order cumulants for modulated symbols. If $s[n]$ represents the nth modulated symbol received amongst a total of N symbols, then cumulants of $s[n]$ can be computed as follows.

(i) Second order cumulants:

$$C_{20} = M_{20}$$
$$C_{21} = M_{21}.$$

(ii) Fourth order cumulants:

$$C_{40} = M_{40} - 3M_{20}^2$$
$$C_{41} = M_{41} - 3M_{20}M_{21}$$
$$C_{41} = M_{41} - M_{20}^2 - M_{21}^2.$$

(iii) Sixth order cumulants:

$$C_{60} = M_{60} - 15M_{20}M_{40} + 30M_{20}^3$$
$$C_{61} = M_{61} - 10M_{20}M_{41} - 5M_{21}M_{40} + 30M_{21}M_{20}^3$$
$$C_{62} = M_{62} - 6M_{20}M_{42} - 8M_{21}M_{41} - M_{22}M40 + 6M_{20}^2M_{22} + 24M_{21}^3M_{20}$$
$$C_{63} = M_{63} - 9M_{21}M_{42} + 12M_{21}^3 - 3M_{20}M42 - 3M_{22}M_{41} + 18M_{20}M_{21}M_{22},$$

where $M_{xy} = \frac{1}{N}\sum_N S^{x-y}\,conj(S)^y$ defines the $(x+y)$th-order moment of the modulated symbols S, and $conj(S)$ is the complex conjugate of S.

SVM-7 [12] uses three fourth-order cumulants, i.e., C40, C41, and C42, and four sixth-order cumulants, i.e., C60, C61, C62, and C63. The cumulant method [12] uses the fourth-order cumulant C40 as the classification statistics.

As shown in Fig. 3, the experiment results show that AlexNet trained with CSI achieves an improved performance over AlexNet trained with dataset as in [12], with an improvement around 10%. Compared to other types of DL-aided AMC schemes, CSI has many advantages:

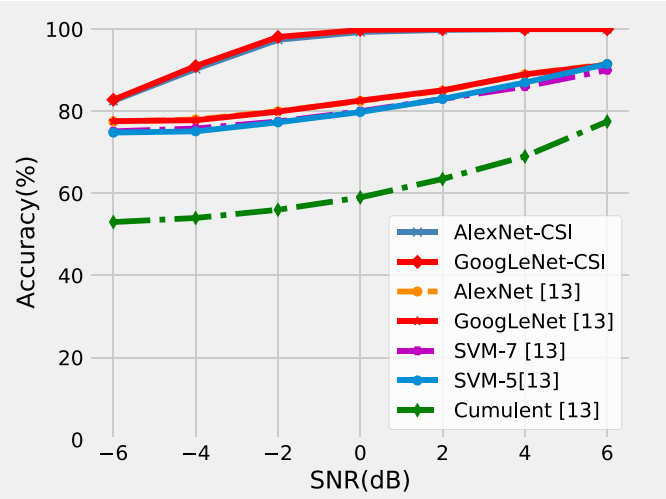


Fig. 3. The average accuracy under different SNR levels for different methods.

- CSI has a strong ability in anti-noise and this power comes from its color distribution based on dot density;
- Due to the separability of CSI, we actually do not need as much training data as in prior work [12], which will reduce the cost of collecting and processing labeled data;
- The colored method based on dot density is a more effective means than the sample distance based method [12].

Compared to shallow classifiers based on expert features, CSI with DL has the following advantages:

- CSI can represent deep-level features, such as amplitude noise, phase error, and so on, in signal waveforms by color distribution;
- CSI is a more effective means for the AMC task than the method proposed in [12];
- Even SVM-7 employs more features, and SVM-5 performs much better than SVM-7 in higher SNR region. It implies that the selection of features for real world modulated signals can be a challenging issue and affect the classification performance. However, DL-based algorithms' self-feature extraction mechanism will require less prior expert knowledge;
- Without expertise in designing new CNN architectures, the existing classical recognition models developed for CV can be directly applied to tackle the AMC task.

B. Application Scenario II: Data Augmentation

It is generally recognized that DL models require a large amount of data in order to achieve good AMC performance. We may have a dataset of images taken in a limited set of conditions in the actual communication scenario. However, insufficient data will lead DL to underestimated parameters and set a wrong decision threshold, which will cause the overfitting problem [33]. One of the possible ways to combat insufficient data is data augmentation [29]. Data Augmentation is routinely used in classification problems. Usually it is non-trivial to encode known invariants in a model. It can be easier to encode those invariants in the data instead of generating additional data items through transformations from existing

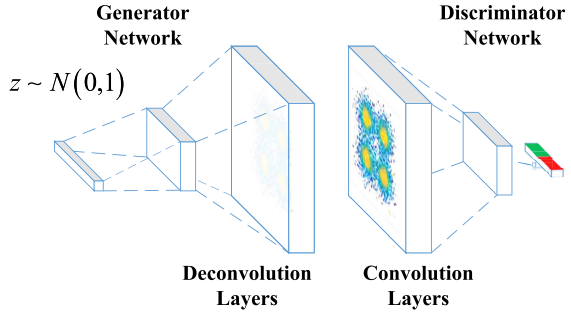


Fig. 4. The GAN model that consists of a generator and a discriminator.

data items. If we know that a class label is invariant to a particular transformation, then we can apply that transformation to generate additional data. If we do not know what transformations might be valid, but we have other data from related problems, we can attempt to learn valid transformations from those related problems and apply them in our setting.

Due to the image data format of CSI, image level data augmentation techniques, including flip, rotation, crop, and so on, can be easily applied. In our prior work [33], we considered image feature level data augmentation, which could be done with GANs. GAN was introduced by Ian GoodFollw in 2014 [29], which can be utilized to generate images via adversarial training. As shown in Fig. 4, there are two basic components, named generator and discriminator, in GAN. The generator takes random noise z as input and produces an image x. The discriminator takes an image x as input and the output is a score that indicates its confidence that the input image is real. The generator's parameters are tuned to fool the discriminator to achieve a high score for the fake images that it generates. The discriminator's parameters are tuned to achieve a high score when its input is a real image, and a low score when the input is a fake image generated by the generator. This model is shown in Fig. 4, where we feed GAN with a CSI image made from OQPSK signals when SNR is 4dB.

In Fig. 4, the generator generates a fake OQPSK image at 4dB SNR, which is very similar to the real OQPSK image at the same SNR level. Both the real image and fake image will be fed into the discriminator. The discriminator will adjust its parameters to distinguish the real image from the CSI dataset from the fake image from the generator. The generator will then tune its parameters to make fake OQPSK images look more like a real CSI. The discriminator and generator are trained adversarially to improve their performance by competing with each other. The traditional GAN loss function in this adversarial game can be defined as:

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim Pdata(x)}[\log D(x)] + \mathbb{E}_{z\sim Pz(z)}[\log(1-D(G(z)))], \quad (2)$$

where $V(G, D)$ is the combined loss of the GAN model, $D(x)$ is the result of the discriminator to identify the real image, $G(z)$ stands for the fake image from the generator, and $D(G(z))$ means the probability of the discriminator to accurately predict a fake image. In the process of generator training, the best result for the generator is $D(G(z)) = 1$, which implies that the

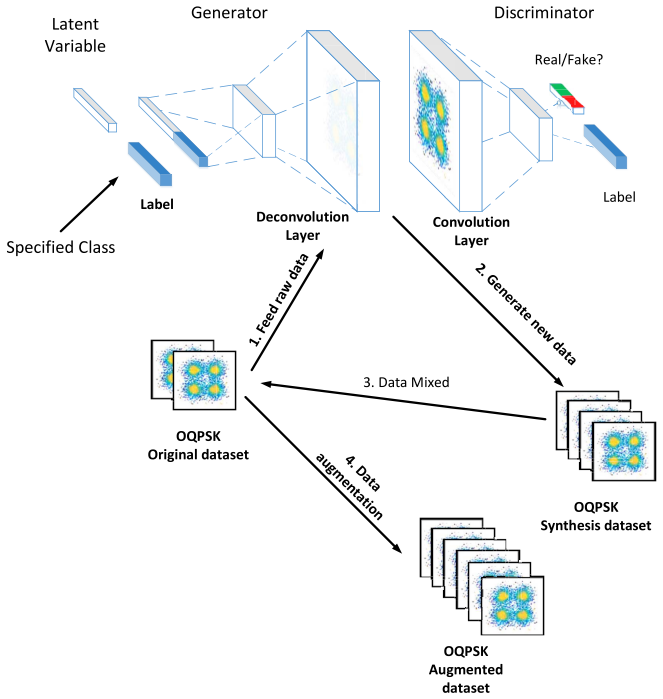


Fig. 5. GAN based data augmentation with CSI processing.

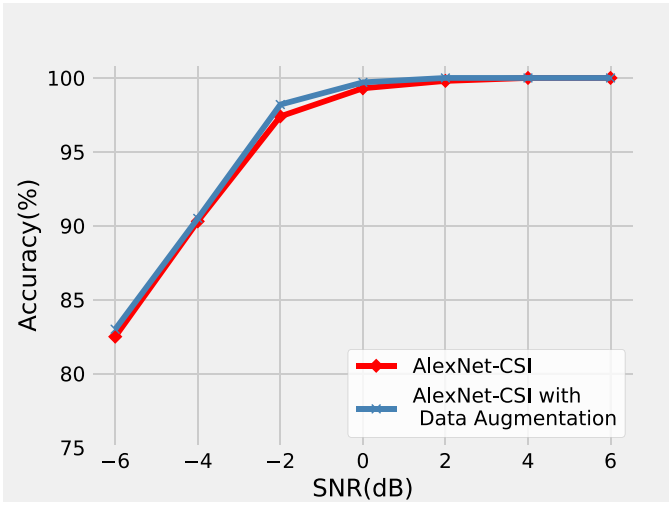


Fig. 6. Comparison results of AlexNet-CSI, AlexNet-CSI with data augmentation.

discriminator has been fooled to predict all the fake images as real.

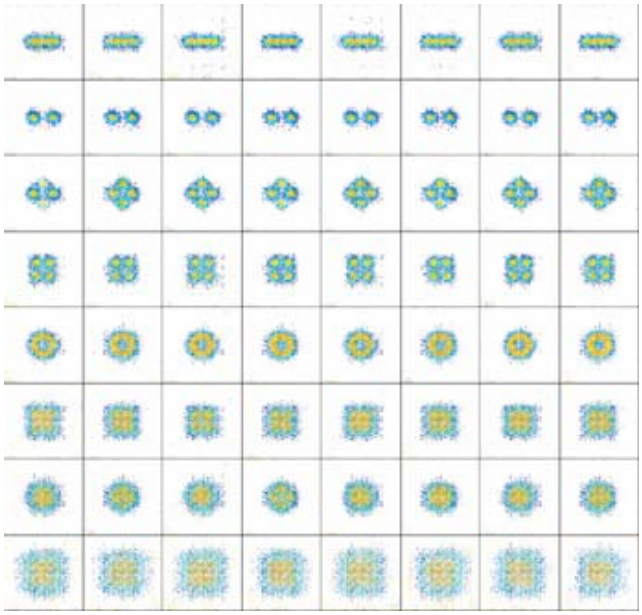
The competition between the generator and the discriminator will stop when the discriminator does not know where the image comes from and the generator cannot improve the quality of generated fake images anymore. Then they reach the stage of a "Nash Equilibrium." A GAN can be utilized to map out a data augmentation manifold. In [33], we choose auxiliary classifier generative adversarial networks (ACGAN) [34] for data augmentation [29]. The discriminator network is trained to figure out whether the sample comes from a real data distribution or a fake data distribution. The classifier in the discriminator network will attempt to make sure the real image comes from the same class. Adversarial training will help the generator network to produce new images from the real images within the same class, but will be represented differently from the training dataset. The proposed scheme is shown in Fig. 5.

Considering the results in Section III-A as a baseline, we use the dataset mentioned in Section III-A to train the ACGAN and to augment the training set. After both the generator and discriminator in the ACGAN converge, we generate 5000 more CSI images to augment the AMC dataset. The results are presented in Fig. 6. From Fig. 6, we can see that the classification accuracy is improved by 0.6%—1.64% over the case without data augmentation. It indicates that: *(i)* Learning a generative manifold by GAN for the classes in the source domain can help to learn better classifiers for the low data target domain; *(ii)* ACGAN is an effective means to conduct CSI-aided AMC data augmentation; *(iii)* The data augmentation technique with ACGAN will slightly improve DNN's classification performance. However, if we want to significantly boost signal classification performance, an effective signal feature extraction method is expected. After 50000 iterations, the losses of discriminator and generator converge, and we obtain the generated images. Having been given the specific label per row, the generator will generate the corresponding images per row. In Fig. 7, we present the sample images created by the generator network.

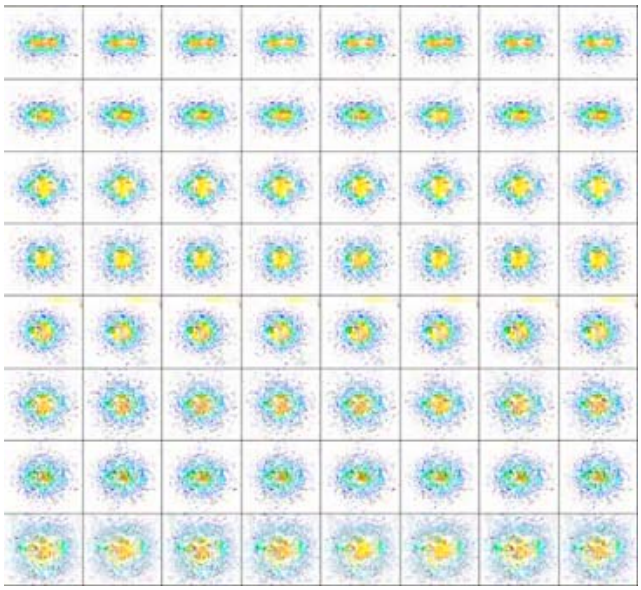
C. Application Scenario III: Specific Emitter Identification

The IoT is such a huge network that contains various sources of emitters. No matter in civilian or military applications, IoT security and confidentiality are of great importance [35], [36]. The specific emitter identification (SEI) technology is to detect the essential difference features from received signal among different communication emitters. The technique identifies different communication radiation sources and is sometimes called "radio frequency fingerprinting." The traditional technique for SEI is to extract signal features using expert knowledge, which may cause unnecessary workload and cannot adapt to time-varying experiments. In this case, we choose to extract specific emitter CSI to learn its deep and imperceptibility features. For example, the CSI size usually indicates the amplitude information of individual communication emitter. The CSI shape could tell the phase information of individual communication emitter. The CSI color distribution also shows distinguishing features. The next step of identification is to feed CSI into a CNN. The dataset could use image data augmentation if there exists data skew or there is insufficient data. The process is depicted in Fig. 8.

To verify the effectiveness of CSI in identifying real emitters, we collect I/Q data from 8 real wireless devices, which emit 16QAM modulated signals. Each emitter CSI consists of 20000 samples of I/Q data. We collected 800 CSI from each emitter for training and 200 for testing. We compare the AlexNet-CSI performance with that of an SVM-aided scheme with five expert features and an SVM-aided scheme with three expert features. The SVM-aided scheme with five expert features uses the following features: Minkowski Bouligand dimension, Marginal spectrum, Covariance distribution, and



(a) SNR = 4dB



(b) SNR = -4dB

Fig. 7. Generated CSI for (a) SNR = 4dB and (b) SNR = −4dB. The rows represent 4ASK, BPSK, QPSK, OQPSK, 8PSK, 16QAM, 32QAM, and 64QAM from top to bottom.

Wavelet entropy feature. The SVM-aided scheme with three expert feature uses Minkowski Bouligand dimension, Marginal spectrum, and Wavelet entropy feature. We extract those expert features and feed it to the classifier. We also consider k-nearest neighbors (KNN) as the classifier and the power spectrum of signals obtained with a fast Fourier transform (FFT) as the expert features, which has 4096, 2048, and 1024 FFT points. Due to balanced testset, we choose the average accuracy as performance measure. After cross-validation, we set the best hyper parameter for SVM and obtain the following result presented in Fig. 9.

As can be seen from Fig. 9, the AlexNet-aided CSI method achieves 100.0% accuracy, the SVM with 5 expert feature

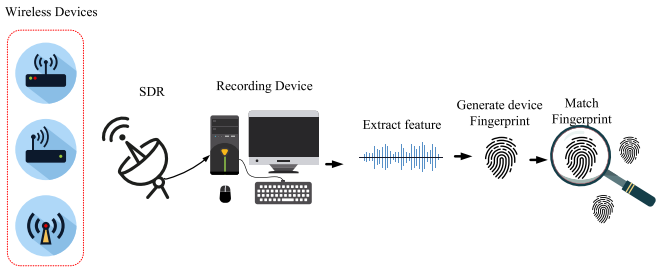


Fig. 8. Specific emitter identification data processing.

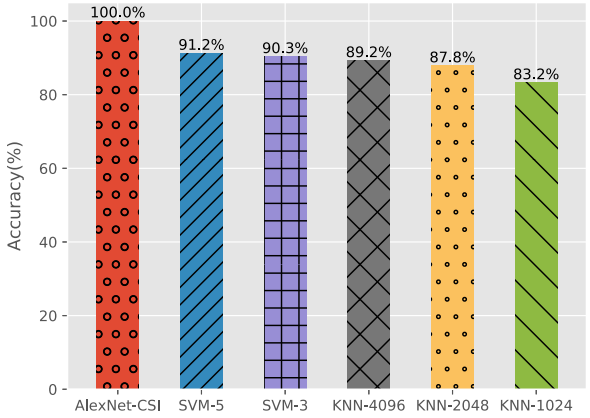


Fig. 9. A comparison study of several specific emitter identification schemes.

scheme achieves 91.2% accuracy, the SVM with 3 expert feature scheme achieves 90.3% accuracy, while the KNN schemes with 4096, 2048, 1024 FFT points achieve 89.2%, 87.8%, and 83.2% accuracy, respectively. The AlexNet-aided CSI method outperforms both the SVM-aided expert feature methods and the KNN-aided expert feature methods. In the real world, the complex and time-varying wireless communication environment will pose great obstacles to precisely utilize expert features. There are usually considerable impairments coming from fading, shadowing, noise, interference, and other types of losses. The AlexNet-aided CSI method is capable of adapting to real signals and learning the emitter features in real-time. This way, the CSI method has a great potential in individual communication emitter recognition.

D. Application Scenario IV: Signal Recognition Semi-Supervised Learning

Most deep learning based classifiers require a large amount of labeled samples for training, but to obtain such labeled data is an expensive and sometimes difficult process. To deal with this limitation, semi-supervised learning provides a good solution, which is a class of techniques that make use of a small amount of labeled data along with a large amount of unlabeled data [37]. Many machine learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvements in learning. GANs have shown a lot of potential in semi-supervised learning where the classifier can obtain a good performance with very few labeled data [38]. In our prior work [13], SSGAN is employed to utilize labeled data

to learn supervision information and unlabeled data to learn data distribution simultaneously. The vanilla architecture of the discriminator has only one output neuron for classifying real and fake (R/F) samples. For the semi-supervised learning, in addition to the R/F neuron, the discriminator will now have N additional neurons for classification of CSIs. We can discard the generator after training, whose only use is to generate unlabeled data to improve the discriminator's performance. Note that the discriminator is turned into an $(N+1)$-class classifier with 1 neuron (i.e., R/F neuron) representing the fake data output and the other N neurons representing real data in N classes. The output is as follows.

- To assert that the R/F neuron output label is "0," when the input is real unsupervised data from the dataset;
- To assert that the R/F neuron output label is "1," when the input is fake unsupervised data from the generator;
- To assert the R/F output label is "0" and the corresponding label output is "1," when the in put is real supervised data.

In other words, the Discriminator has three different sources of training data as:

- Real images with labels: these are image-label pairs like in any regular supervised classification problem;
- Real images without labels: for these, the classifier only learns that these images are real;
- Images from the generator: for these, the discriminator learns to classify them as fake.

The combination of these different sources of data will make the classifier learn from a broader perspective. The process is illustrated in Fig. 10. Feature matching changes the cost function of the generator to minimizing the statistical difference between the features of the real data and the generated data. Often, we measure the L2-distance between the means of their feature vectors, as

$$\min \left\| \mathbb{E}_{x\sim p_{\text{data}}} M(x) - \mathbb{E}_{z\sim p_z(z)} M(G(z)) \right\|, \qquad (3)$$

where $M(x)$ is the feature vector extracted in an immediate layer by the discriminator. Therefore, feature matching expands the goal from beating the opponent to matching features in real data.

For semi-supervised learning, we shrink the datasets proposed in Section III-A to 500, 1000, 2000, and 5000 of labeled data samples for each modulation category, meaning we have 4000, 8000, 16000, and 40000 labeled data samples for the overall dataset, respectively. For semi-supervised learning, we select SSGAN to be our classification model and set SSGAN's ImageNet hyper parameter as that used in [34]. We just train the SSGAN discriminator D without updating the generator G as in the CNN baseline scheme. The classification accuracy results are presented in Table I. From Tab. I, we can make the following observations.

- The amount of labeled data plays an important role in semi-supervised learning: more labeled data help to improve the classifier's accuracy performance.
- SSGAN, which shares the same classifier architecture as the baseline scheme but only uses a small labeled dataset, outperforms CNN when the dataset is small, since it utilizes three sources of data, i.e., real labeled data, real unlabeled data, and fake unlabeled data, instead of only one data source for training as in the case of CNN. With a larger labeled dataset, the unavoidable noise in the GAN generated data will prevent the SSGAN classifier to learn the correct features.
- GANs can be considered as a potential solution for learning complex tasks with limited labeled data, and CSI can bridge the gap between signal recognition and SSGAN.

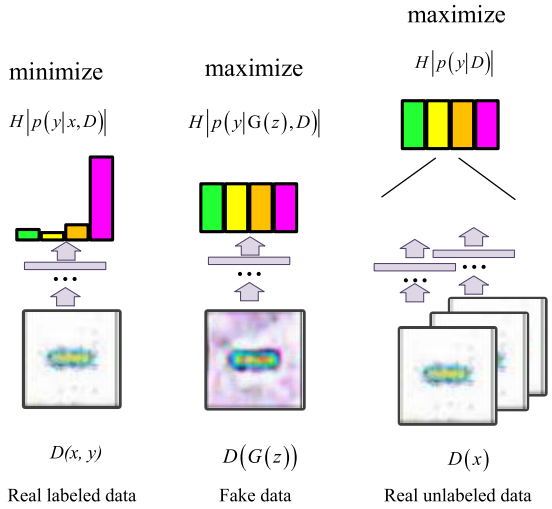


Fig. 10. A sketch of the CSI information flow through the generator and discriminator neural networks for semi-supervised learning.

TABLE I
CLASSIFICATION ACCURACY OF CNN AND SSGAN

Labeled Data	Baseline (%)	SSGAN (%)
4000	64.89	66.12
8000	72.86	72.99
16000	88.45	87.31
40000	97.32	93.44

E. Application Scenario V: DNN Deployment

As mentioned before, CSI has many potential application scenarios in signal recognition. However, the computational power and memory requirements pose potential obstacles to deploy DNN at an edge device. Although we could offload some computation tasks to the cloud, cloud computing requirements for bandwidth, latency, and availability still present severe challenges to real-time tasks. In such cases, local execution is still needed. As a remedy, the neural network compression technique provides a promising way to pursuit the balance between computation cost and accuracy [40].

In Sections III-A and III-C, CSI exhibits high promise as a new form of data, where a high-performance CNN can first be trained to achieve high accuracy in AMC or communication emitter recognition. We hope a CNN trained with CSI data can be compressed by a large factor to make it possible to deploy the compressed CNN at edge devices with only slightly sacrificed accuracy. When designing the CNN architecture, CV experts often arrive at the design once they think the network has enough representation power for the ImageNet data distribution. However, when it comes to CSI

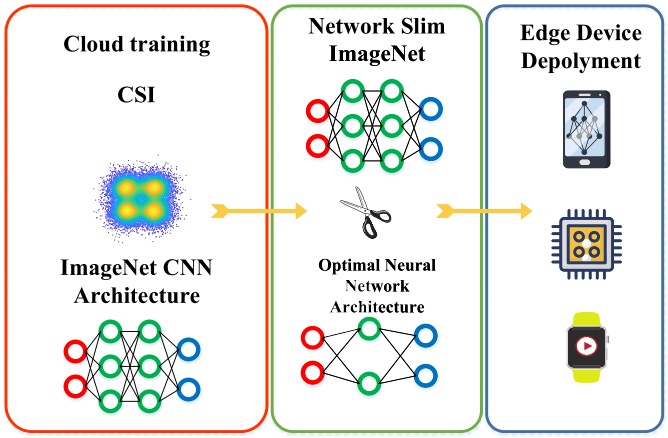


Fig. 11. The network compression techniques applied to the AlexNet-CSI method.

for signal recognition, such empirically designed networks are usually heavily oversized. We could use the network compression technique to remove much redundancy in the DNN model before deploying it at an edge device. Neural network compression techniques include parameter pruning and sharing, low rank factorization, transferred/compact convolutional filters, and knowledge distillation. We will choose one of these neural network compression techniques to compress the network.

In our prior work [14], we pretrain AlexNet with CSI and choose the average percentage of zeros (APoZ) criterion [13] to prune the network. APoZ is a heuristic criterion to evaluate the importance of each filter. This criterion calculates the sparsity of each channel in output activations as its importance score, computed as:

$$s_i = \frac{1}{|\alpha(i,:,:))|} \sum\sum I(\alpha(i,:,:)) == 0), \tag{4}$$

where $|\alpha(i,:,:))|$ is the number of elements in the ith channel of tensor α after ReLU activation, and $I(\cdot)$ denotes the indicator function.

The dataset we choose is the same as in Section III-A. We evaluate the performance of the slimming process from classification accuracy, floating point operations (FLOPs), parameter and actual edge device validation perspectives. Compared to the original CNN, the experiment results show that the light CNN convolution layer could have only 1.5% $\sim$ 5% of parameters and use 33% $\sim$ 35% of time, with no more than 1% degradation of accuracy at every SNR level.

1) Accuracy Comparison: In this experiment, we compare the recognition accuracy of (i) the pruned neural network under the ensemble learning rules with the voting method, (ii) the pruned neural network without ensemble learning, and (iii) the original neural network under different SNR levels. The results are shown in Fig. 13. The average accuracy is computed as:

$$\text{Accuracy} = \frac{N_{correct}}{N} \times 100\%, \tag{5}$$

where $N_{correct}$ is the number of correctly classified test samples and N is the total number of test samples.

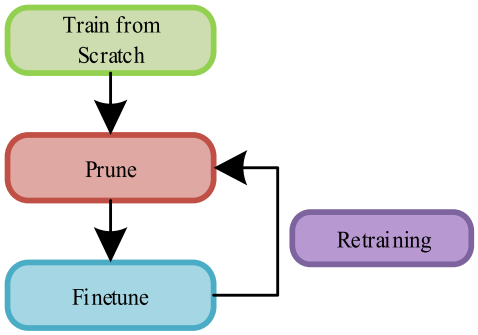


Fig. 12. Illustrating the pruning process.

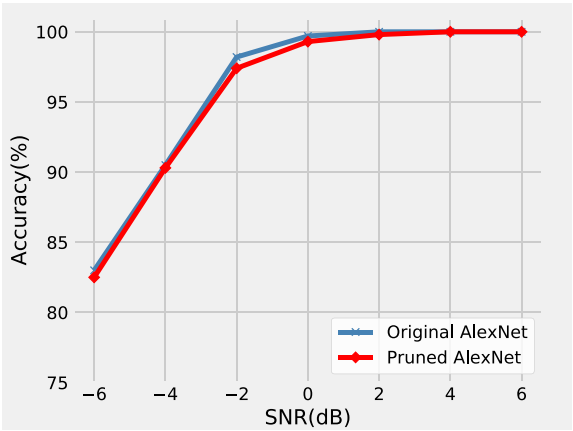


Fig. 13. Accuracy comparison among the original AlexNet and the pruned AlexNet neural network.

The pruning algorithm compresses the neural network structure, which inevitably causes loss of the recognition accuracy. The recognition accuracy of a neural network after pruning is usually decreased by about 0.5% $\sim$ 1% [13]. The results in Fig. 13 show that APoZ is a valid criterion to prevent pruned neural network from significant accuracy loss.

2) FLOPs Comparison: FLOPs is an evaluation criterion usually used to measure the complexity of a model, which has a crucial influence on the reasoning speed of the model. In CNNs, the convolution layer contributes to most of the FLOPs. In this article, we take the ratio of the FLOPs of the original model to the pruned model, denoted by FCR, as an evaluation criterion. The FCR is defined as:

$$FCR = \frac{FLOPs(orignial)}{FLOPs(pruned)}. \tag{6}$$

As can be observed in Fig. 14, the neural network pruning method based on APoZ can achieve a certain compression effect under each SNR level. Furthermore, the compression effect at high SNRs (i.e., 0dB $\sim$ 6dB) is much more obvious than that of low SNRs (i.e., $-$6dB $\sim$ $-$2dB). This is due to the fact that the data distribution of the CSI is simpler under a high SNR and there is more redundancy in the neural network while ensuring a certain recognition accuracy.

3) Number of Parameters Comparison: The number of parameters is an evaluation criterion used to measure the storage requirement of the model. In this article, the compression ratio of parameters (PCR) before and after pruning is also used

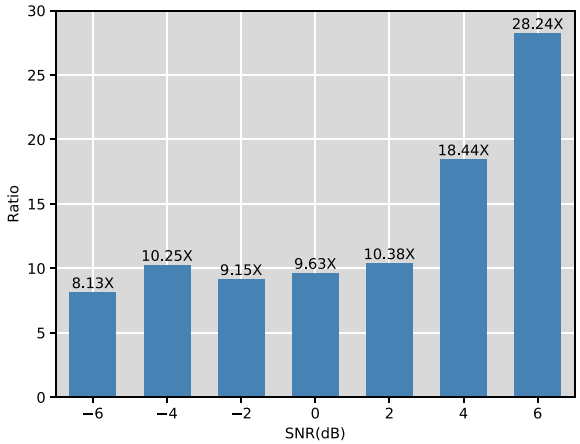


Fig. 14. FCR results at different SNR levels.

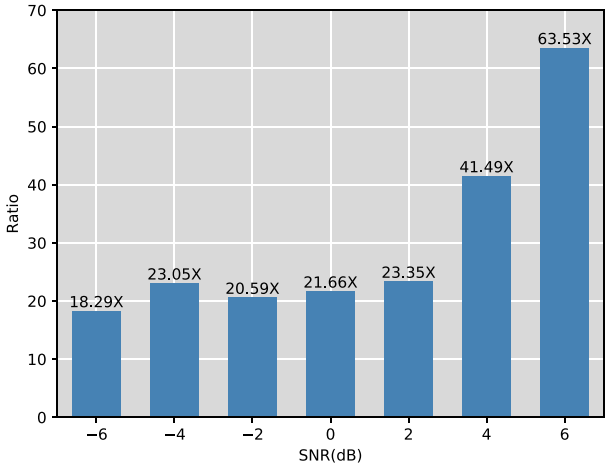


Fig. 15. PCR results at different SNR levels.

Fig. 16. NVIDIA Jetson TX2 Module.

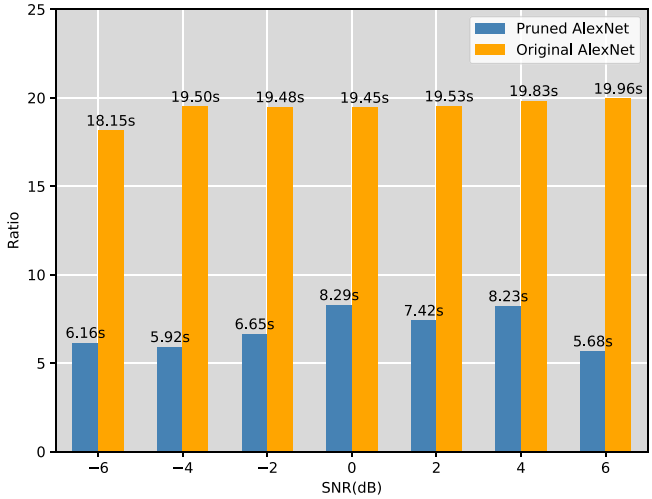


Fig. 17. Prediction time comparison for 1000 samples.

as a metric to measure the pruning effect of neural networks, which is defined as:

$$PCR = \frac{\text{Number of Parameters(orignial)}}{\text{Number of Parameters(pruned)}}. \tag{7}$$

The results are shown in Fig. 15. First, the neural network pruning method based on APoZ can achieve a good PCR compression effect for each SNR level, which is similar to the FLOPs comparison. The reason for this effect is that there is a lot of redundancy in the fully connected layer of the neural network. In addition, the compression effect of high SNRs (i.e., 0dB $\sim$ 6dB) is much more obvious than that of low SNRs (i.e., $-$6dB $\sim$ $-$2dB). This is because the data distribution of CSI is simpler in the high SNR region, and thus there is more room for compression.

4) Device Validation: In order to verify the application value of our proposed scheme on actual edge devices, we deploy the original AlexNet and the pruned AlexNet into real edge devices, i.e., a NVIDIA Jetson TX2 with its GPU acceleration turned off (see Fig. 16), and compare their performance.

To reduce the random factor, we have tested the reasoning time of 1000 CSI of the two models in each experiment and obtained the average results of 50 experiments. The results are shown in Fig. 17. It can be seen that the pruned AlexNet at all SNR levels is about 3 times faster in predicting samples in the NVIDIA Jetson TX2 Module, since we cut down the amount of FLOPs in the convolution layer. We can also see that the prediction times are almost equal at every SNR level. Even at 6dB, the pruned AlexNet' FLOPs are much smaller than the other AlexNet's FLOPs at other SNR levels. We believe this is caused by the communication delay. From the above experimental results, it can be seen that we significantly slim the network for reduced time and storage requirements, but without losing more than 1% accuracy at every SNR level.

IV. Future Research Opportunities

As mentioned above, CSI has strong capabilities to deal with the challenging signal recognition problem, and it can effectively improve the performance of many applications such as AMC, data augmentation, individual communication emitter recognition, and neural network compression. Moreover, CSI opens a door for many more research opportunities. In this section, we introduce two of them as examples: *(i)* CSI-aided transfer learning and *(ii)* signal recognition dataset annotation.

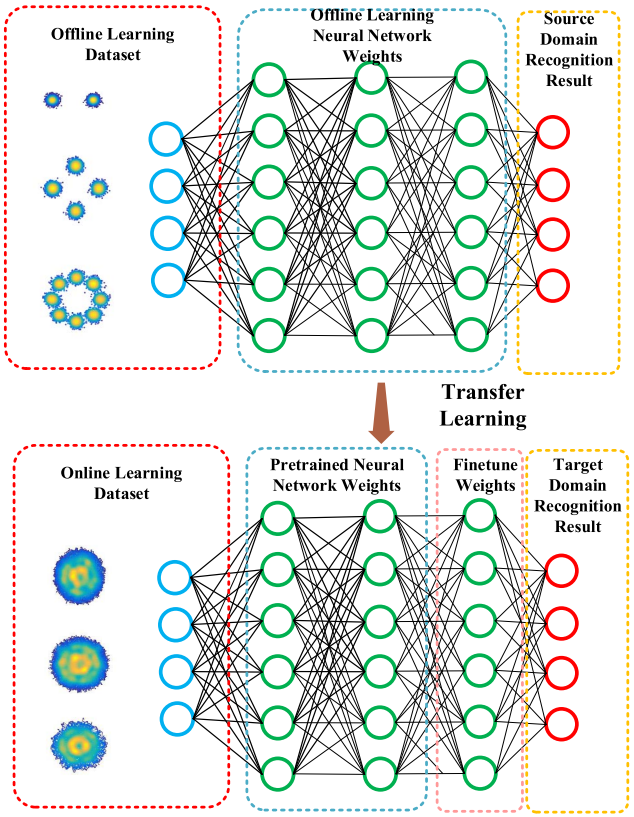


Fig. 18. CSI-aided transfer learning process.

A. Transfer Learning

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a different task [41]. It is currently highly popular in the field of DL because it enables us to train a DNN with comparatively small amount of data. The real wireless communication environment is complex and heterogeneous [8]. For some problems where we may not be able to obtain sufficient data, transfer learning will enable us to fully exploit the features extracted from the source data. Transfer learning will also help us to make full use of the high-speed incoming signal data before it expires [9]. For the CSI dataset, we propose a possible transfer learning scheme with DNN, which consists of two stages including offline learning and online learning.

1) Offline Learning: As shown in Fig. 18, we select a related predictive problem with abundant data where there is some relationship in the input data, output data, and the features learned during the mapping from input to output data. We should make sure that the model be better than a naive model to ensure that some feature learning has been performed. After offline learning, the DNN will learn wireless communication signal CSI with useful low-level and generalizable features, which will be the starting point to the next stage.

2) Online Learning: The model fitted on the original task can then be used as the starting point for a model on a different task [42]–[44]. Optionally, the model may need to be adapted or refined on the input-output pair data available for the next task. If we have a small CSI dataset for the second task, we can freeze the weights in most of the layers to avoid overfitting. As is shown in Fig. 18, if we get a large CSI dataset for the second task, we should freeze the weights in fewer layers to avoid underfitting. Compared to traditional machine learning, transfer learning with CSI will help wireless communication environment analysis to be more adaptive. Further, we can also utilize a lot of data from simulations and long-term accumulation, which will free us from the problem of lack of data. In extremely data-limited scenarios, transfer learning can also conduct few-shot or one-shot learning [45] due to CSI's superior feature extraction power. Moreover, transfer learning will accelerate the training process when the former training checkpoint is used as the start point.

B. Signal Recognition Dataset Annotation

Most successful deep learning techniques require ground-truth labels for a sufficient training dataset. Nevertheless, in many tasks, it is difficult to attain strong supervision information due to the high cost of data labelling process [46]. Wireless communication is not an exception. It is desired for machine learning techniques to be able to annotate wireless dataset with weak supervision information. This situation is often called weakly supervised learning and can be divided into three types: incomplete supervision, inexact supervision, and inaccurate supervision. The incomplete supervision task generally refers to semi-supervised learning [46], which has been discussed in Section III. So we will focus on the discussions of CSI applications in inexact supervision and inaccurate supervision tasks in the following.

1) Inexact Supervision: Inexact supervision concerns about the situation where some wireless communication data supervision information is given, but not as exact as desired. As a remedy, we can use multi-instance learning with CSI to address the inexact supervision issues. The original goal of multi-instance learning is to predict labels for unseen signal attributes. For example, we are not satisfied with just knowing modulation type of the signal waveform, but we also want to know which signal emitter generated the signal. Multi-instance learning with CSI is an effective alternative to achieve this dual-goal.

2) Inaccurate Supervision: Inaccurate supervision concerns about the situation where the supervision information is not always the ground-truth. In other words, some label information may suffer from errors. When annotating abundant unlabeled dataset, crowd sourcing is commonly used as a cost-effective way to collect labels for training data. Specifically, unlabeled instances are outsourced to a large group of workers to label them. With the increase of work, annotation accuracy will decrease if the worker is incapable of annotating some samples. However, CSI has good separability and anti-noise merit as discussed above. The annotate workers will be more confident about CSI-based datasets and will produce a labeled dataset with higher quality in turn.

In summary, the two different situations are illustrated in Fig. 19. In the inexact supervision scenario, we also want to know more wireless signal fine-grained attributes, such as SNR, transmitter, and phase offset. CSI will reveal these fine-grained features via its shape and color. In the inaccurate supervision scenario, just as discussed in Sections III-A

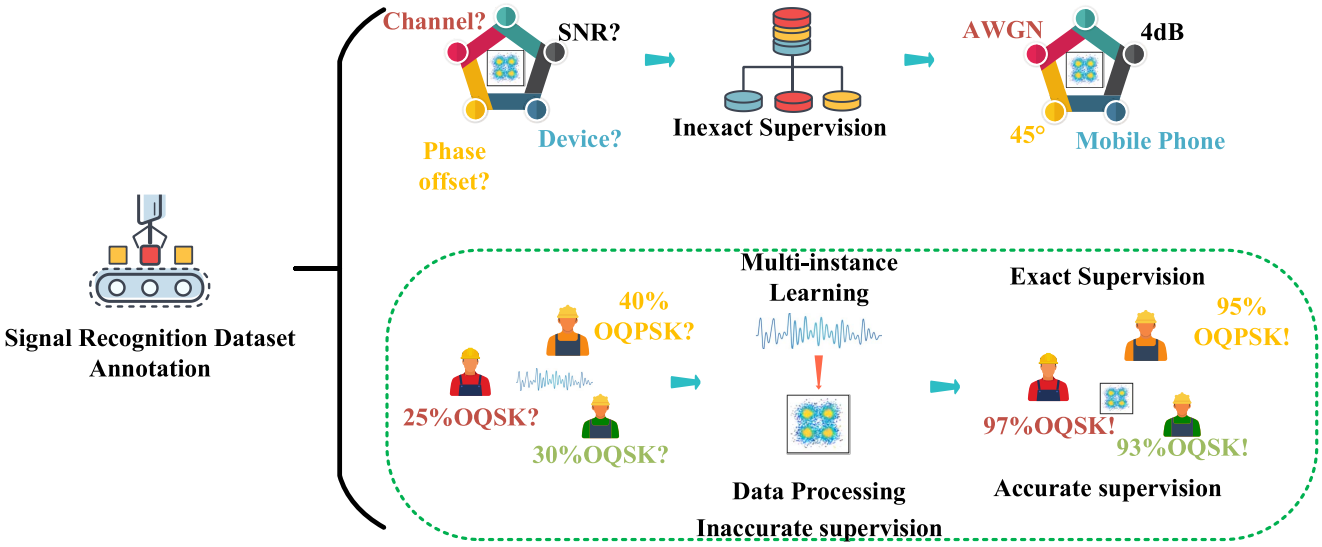


Fig. 19. The signal recognition dataset annotation process.

and III-C, CSI is a more effective method to learn wireless signal's hidden features and will help data annotation workers to make more accurate decisions.

V. Conclusion

Nowadays, DL has attracted much attention in various fields and has great potential to meet the requirements of emerging signal recognition scenarios. However, there still exists a big gap between signal recognition and DL models, since the structure of signal data is very different from that for which DL has been proven successful. In this article, we proposed CSI, a bridge from wireless signals to DL models. From the part of signal processing, CSI could serve as a 3-dimensional eye pattern. The received wireless signal features are well retained in the CSI of the signal. Some types of distortion, such as Gaussian noise, non-coherent single frequency interference, phase noise, amplifier compression, etc., can be learned as imperceptibility features. In other words, CSI has the potential to represent the signal or individual communication emitter, which can provide both in-phase and quadrature information. From the perspective of DL, CSI could be treated as a general image data format. Quite a few practical DL methods, such as data argumentation, signal recognition dataset annotation, transfer learning, and deep neural network compress techniques can be applied to CSI-aided DL. We presented several application scenarios where CSI has been successfully applied with superior performance demonstrated, and concluded this article with a discussion of several future research opportunities.

References

[1] J. Zhu, C. Gong, S. Zhang, M. Zhao, and W. Zhou, "Foundation study on wireless big data: Concept, mining, learning and practices," *IEEE/CIC China Commun.*, vol. 15, no. 12, pp. 1–15, Dec. 2018.

[2] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP J. Adv. Signal Process.*, vol. 2016, p. 67, May 2016.

[3] K. Zheng *et al.*, "Big data-driven optimization for mobile networks toward 5G," *IEEE Netw.*, vol. 30, no. 1, pp. 44–51, Jan. 2016.

[4] C. Clancy, J. Hecker, E. Stuntebeck, and T. J. O'Shea, "Applications of machine learning to cognitive radio networks," *IEEE Wireless Commun.*, vol. 14, no. 4, pp. 47–52, Aug. 2007.

[5] C. Jiang, N. Beaulieu, L. Zhang, Y. Ren, M. Peng, and H. Chen, "Cognitive radio networks with asynchronous spectrum sensing and access," *IEEE Netw.*, vol. 29, no. 3, pp. 88–95, Jun. 2015.

[6] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen, and L. Han, "Machine learning paradigms for next generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[7] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.

[8] T. J. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[9] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1389–1401, Jun. 2019.

[10] G. Gui, M. Liu, F. Tang, N. Kato, and F. Adachi, "6G: Opening new horizons for integration of comfort, security and intelligence," *IEEE Wireless Commun. Magn.*, early access, Mar. 3, 2020, doi: 10.1109/MWC.001.1900516.

[11] R. Li *et al.*, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 175–183, Oct. 2017.

[12] S. Peng *et al.*, "Modulation classification based on signal constellation diagrams and deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 99, pp. 1–10, Mar. 2018.

[13] Y. Tu, Y. Lin, J. Wang, and J. U. Kim, "Semi-supervised learning with Generative Adversarial Networks on digital modulation classification," *Comput. Mater.*, vol. 55, no. 2, pp. 243–254, May 2018.

[14] Y. Tu and Y. Lin, "Deep neural network compression technique towards efficient digital signal modulation recognition in edge devices," *IEEE Access*, vol. 7, pp. 58113–58119, 2019.

[15] J. Tian, Y. Pei, Y. Huang, and Y. Liang, "Reinforcement learning based capacity management in multi-layer satellite networks," *IEEE Trans. Commun.*, vol. 4, no. 4, pp. 894–907, Dec. 2018.

[16] C. Jiang and X. Zhu, "Modulation-constrained clustering approach to blind modulation classification for MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4685–4699, Jul. 2020.

[17] G. Li, A. Hu, J. Zhang, L. Peng, C. Sun, and D. Cao, " High-agreement uncorrelated secret key generation based on principal component analysis preprocessing," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 3022–3034, Jul. 2018.

[18] J. Yu, A. Hu, G. Li, and L. Peng, "A robust RF fingerprinting approach using multi-sampling convolutional neural network," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6786–6799, Aug. 2019.

[19] R. Wu, X. Chen, H. Han, H. Zhao, and Y. Lin, "Abnormal information identification and elimination in cognitive networks," *Int. J. Performability Eng.*, vol. 14, no. 10, pp. 2271–2279, Oct. 2018.

[20] G. Ding, Q. Wu, L. Zhang, Y. Lin, T. Theodoros, and D. Yao, "An amateur drone surveillance system based on the cognitive Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 29–35, Jan. 2018.
[21] Y. Lin, X. Zhu, Z. Zheng, Z. Dou, and R. Zhou, "The individual identification method of wireless device based on dimensionality reduction and machine learning," *J. Supercomput.*, vol. 75, no. 5, pp. 3010–3027, Jun. 2019.
[22] Y. Lin, C. Wang, J. Wang, and Z. Dou, "A novel dynamic spectrum access framework based on reinforcement learning for cognitive radio sensor networks," *Sensors*, vol. 16, no. 10, p. 1675, Oct. 2016.
[23] T. Liu, Y. Guan, and Y. Lin, "Research on modulation recognition with ensemble learning," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, p. 179, Dec. 2017.
[24] Z. Zhang, X. Guo, and Y. Lin, " Trust management method of D2D communication based on RF fingerprint identification," *IEEE Access*, vol. 6, pp. 66082–66087, 2018.
[25] J. Wang, S. Guan, C. Jiang, D. Alanis, Y. Ren, and L. Hanzo, "Network association in machine-learning aided cognitive radar and communication co-design," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2322–2336, Oct. 2019.
[26] X. Zhang, T. Y. Seyfi, S. Ju, S. Ramjee, A. E. Gamal, and Y. C. Eldar, "Deep learning for interference identification: Band, training SNR, and sample selection," in *Proc. IEEE SPAWC*, Cannes, France, Aug. 2019, pp. 1–5 .
[27] I. Dagres *et al.*, "An integrated platform for source detection, identification and localization with applications to cognitive-radio," in *Proc. Eur. Wireless*, Guildford, U.K., Apr. 2013, pp. 1–6.
[28] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 541–551, Apr. 2012.
[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[30] K. Simonyan and A. Zisserman. (Apr. 2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. [Online] Available: https://arxiv.org/abs/1409.1556.
[31] Y. Wang, J. Yang, M. Liu, and G. Gui, "LightAMC: Lightweight automatic modulation classification using deep learning and compressive sensing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3491–3495, Mar. 2020.
[32] F. Meng, P. Chen, L. Wu, and X. Wang, "Automatic modulation classification: A deep learning enabled approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10760–10772, Nov. 2018.
[33] B. Tang, Y. Tu, Z. Zhang, and Y. Lin, "Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks," *IEEE Access*, vol. 6, pp. 15713–15722, 2018.
[34] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. ICML*, Sydney, NSW, Australia, Aug. 2017, pp. 2642–2651.
[35] A. C. Polak and D. L. Goeckel, "Identification of wireless devices of users who actively fake their RF fingerprints with artificial data distortion," *IEEE Wireless Commun.*, vol. 14, no. 11, pp. 5889–5899, Nov. 2015.
[36] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios," *IEEE Wireless Commun.*, vol. 23, no. 5, pp. 60–67, Nov. 2016.
[37] T. J. O'Shea, N. West, M. Vondal, and T. C. Clancy, "Semi-supervised radio signal identification," in *Proc. ICACT*, PyeongChang, South Korea, Feb. 2017, pp. 33–38.
[38] J. T. Springenberg, "Unsupervised and semi-supervised learning with categorical generative adversarial networks," in *Proc. ICLR*, San Diego, CA, USA, May 2015, pp. 1–20.
[39] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. "Improved techniques for training GANs," in *Proc. NIPS*, Barcelona, Spain, Dec. 2016, pp. 2234–2242.
[40] J. Luo, H. Zhang, H. Zhou, C. Xie, J. Wu, and W. Lin, "ThiNet: Pruning CNN filters for a thinner net," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2525–2538, Oct. 2019.
[41] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proc. ICML Workshop Unsupervised Transfer Learn.*, Edinburgh, U.K., Jun. 2012, pp. 17–36.
[42] J. T. Zhou, H. Zhang, D. Jin, and X. Peng, "Dual adversarial neural transfer for sequence labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jul. 29, 2019, doi: 10.1109/TPAMI.2019.2931569.
[43] J. T. Zhou, I. Tsang, S. Pan, and M. Tan, "Multi-class heterogeneous domain adaptation," *J. Mach. Learn. Res.*, vol. 20, pp. 1–31, Jan. 2019.
[44] J. T. Zhou, S. J. Pan, and I. W. Tsang, "A deep learning framework for hybrid heterogeneous transfer learning," *Artif. Intell. Rev.*, vol. 275, pp. 310–328, Jun. 2019.
[45] F. Li, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006.
[46] Z. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Jan. 2018.

Yun Lin (Member, IEEE) received the B.S. degree from Dalian Maritime University, Dalian, China, in 2003, the M.S. degree from the Harbin Institute of Technology, Harbin, China, in 2005, and the Ph.D. degree from Harbin Engineering University, Harbin, in 2010, where he is currently a Full Professor with the College of Information and Communication Engineering. He was a Research Scholar with Wright State University, USA, from 2014 to 2015. His current research interests include machine learning and data analytics over wireless networks, signal processing and analysis, cognitive radio and software defined radio, artificial intelligence, and pattern recognition. He is serving as an Editor for the IEEE TRANSACTIONS ON RELIABILITY, *KSII Transactions on Internet and Information Systems*, and *International Journal of Performability Engineering*. In addition, he served as the General Chair of ADHIP 2020, the TPC Chair of MOBIMEDIA 2020, ICEICT 2019, and ADHIP 2017, and the TPC Member of GLOBECOM, ICC, ICNC, and VTC.

Ya Tu (Student Member, IEEE) received the B.S. degree from the College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, China, in 2016. He is currently pursuing the Ph.D. degree with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China. His current research interests include signal processing, machine learning, and data analysis.

Zheng Dou (Member, IEEE) received the B.S. degree from Harbin Engineering University, China, in 2001, and the Ph.D. degree from the College of Information and Communication Engineering, Harbin Engineering University, in 2007, where he is currently a Professor. His research interests include intelligent wireless communication, cognitive radio, and the design of communication signal waveform.

Lei Chen (Senior Member, IEEE) received the B.Eng. degree in computer science and applications from the Nanjing University of Technology, China, in 2000, and the Ph.D. degree in computer science and software engineering from Auburn University, USA, in 2007. He is currently an Associate Professor and the Graduate Program Director for the Department of Information Technology, Georgia Southern University, USA. His research interests focus on the security and privacy of networks, information, cloud, big data, mobile, handheld, and wireless networks, and digital forensics.

Shiwen Mao (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from Polytechnic University, Brooklyn, NY, USA, in 2004. He is a Professor, the Earle C. Williams Eminent Scholar, and the Director of Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. His research interests include wireless networks, multimedia communications, and smart grid. He was a recipient of the IEEE VTS 2020 Jack Neubauer Memorial Award, the IEEE ComSoc TC-CSR Distinguished Technical Achievement Award in 2019, the NSF CAREER Award in 2010, and the 2004 IEEE ComSoc Leonard G. Abraham Prize in the Field of Communications Systems.